WPS3496

What Factors Influence World Literacy?

Is Africa Different?

By

Dorte Verner[1]

World Bank Policy Research Working Paper 3496, January 2005

[1] I would like to thank Helena Ribe for invaluable support and Niels-Hugo Blunch for research assistance..

Abstract

Ninety-five percent of the world's illiterate people live in developing countries, and about 70 percent are women. Female illiteracy rates are particularly high in Sub-Saharan Africa. In Niger and Burkina Faso, for example, more than 90 percent of women are illiterate. This paper presents a model of literacy. It shows that the main determinants of worldwide literacy are enrollment rates, average years of schooling of adults, and life expectancy at birth. Income has a weak nonlinear effect, negatively affecting literacy until a threshold level of per-capita income of about $2,200 a year is reached and positively affecting literacy thereafter. Finally, African countries do not have a significantly higher illiteracy rate when controlling for other factors.

1. Introduction

Nearly a billion people, two-thirds of them women, entered the twenty-first century unable to read a book or write their names. Their illiteracy will limit their earning power and affect a range of other social and economic variables. Identifying the determinants of illiteracy is crucial if the problem and its consequences are to be ameliorated. This paper is an attempt to shed light on these issues in a world context.[2]

Ninety-five percent of the world's illiterate people live in developing countries, and about 70 percent are women. Female illiteracy rates are particularly high in Sub-Saharan Africa. In Niger and Burkina Faso, for example, more than 90 percent of women are illiterate (EdStats, World Bank).

Together with an array of other indicators, the adult literacy rate has been used to assess the level of a country's development. It is also used as an output indicator of the quality and effectiveness of school systems. The most commonly used definition of a literate person, that of UNESCO, is one "who can engage in all those activities in which literacy is required for effective functioning of his/her reference group and community and also for enabling him/her to continue to use reading, writing and calculation for his/her own and the community's development" (UNESCO 1993). Unfortunately, most of the data on literacy in developing countries are on literacy, not functional literacy. This study is therefore based on literacy data.

[2] Blunch and Verner take a case study approach to the same problem. Their paper shows that literacy and income are positively correlated in Ghana.

The paper is organized as follows. Section 2 takes an historical look at literacy and examines the benefits of widespread literacy. Section 3 examines literacy rates in different countries. Section 4 presents the model, data, and methodology. Section 5 presents findings and Section 6 summarizes the study's findings.

2. The Development and Benefits of Widespread Literacy

Large-scale literacy is a recent phenomenon. Before the nineteenth century, public schools did not exist, and education was reserved largely for the elite. In many industrial countries, a high rate of literacy developed because Protestants wanted to be able to read the Bible. Their demands for literacy training eventually led to agitation for public education during the nineteenth century (Hass 1996).

Public school systems continued to develop in the twentieth century, largely because of increased and more specialized industrialization. As economic growth increased as a result of industrialization, demands for a more educated and specialized labor force grew and literacy increased.

Proponents of literacy programs argue that reading and writing skills improve individuals' chances of increasing their earnings. While it is difficult to determine causality, a growing body of evidence suggests that literacy increases the productivity and earning potential of a population. In Thailand, for example, farmers with four years of schooling are three times more likely to adopt new fertilizers and other inputs than

farmers with one to three years of schooling (World Bank 1991). The difference appears to be attributable to the fact that better educated farmers absorb new information more quickly and are more innovative than less educated farmers.

Gains of a nonmarket nature may also accrue from literacy. Literate people tend to be more aware of health and nutrition issues, for example. Literate women also tend to have fewer children than illiterate women. Literacy is thus associated with decreasing fertility and mortality rates (see Cochrane 1979 and Verner 1995). Literacy also leads to greater literacy in the next generation, as literate parents are more likely than illiterate parents to understand the importance of education.

Illiteracy is caused by several factors, including absenteeism, repetition, and dropping out of school. Low school enrollment rates are caused by both in-school factors, such as the availability, quality, and efficiency of schools, and out-of-school factors, such as direct and opportunity costs.

Investing in adult literacy programs yields economic, social, and political benefits (Valerio 1997). Increased adult literacy enhances the human capital stock, increasing economic growth rates and improving social indicators. It is an important determinant of income (Neumann and Weiss 1995) and is positively associated with other social indicators, such as reduced infant and child mortality rates and improved child nutrition. In the United States it is inversely correlated with the costs associated with

unemployment, incarceration, criminal proceedings, and law enforcement (Valerio 1997). Literacy is also a means of enhancing national, social, and political solidarity.

Little has been written about the determinants of literacy. In one of the few empirical studies, Levy, Spratt, and Leboucher (1995) analyze self-reports of literacy among 9- to 69-year-olds in Morocco. They find that illiteracy rates have fallen by about 50 percent over the past 30 years. Women tend to be less literate than men, however, and the urban-rural gap in literacy has increased. Regression analysis indicates that parents' literacy, urban environment, and the quintile of household expenditure are all positively associated with literacy.

3. Literacy Rates across Regions

The number of illiterate people in the world is estimated at 962 million as of 1990 (EdStats, World Bank). Although the rate of illiteracy has fallen—from 44 percent of the world's population in 1950 to about 25 percent in 1990—the number of illiterate people is rising, as a result of high population growth rates, inadequate schooling, low school enrollment rates, and poverty.

Ninety-five percent of the world's illiterate people live in developing countries, and about 70 percent are women. Female illiteracy rates are particularly high in Sub-Saharan Africa. In Niger and Burkina Faso, for example, more than 90 percent of women are illiterate (EdStats, World Bank). Moreover, throughout many regions of the world the absolute number of illiterate people is rising, as a result of high population growth rates, the low quality of schooling, low school enrollment rates, and poverty.

Illiteracy rates vary with income levels and across regions (table 1). In low-income countries, illiteracy is almost three times as high as in lower-middle and upper-middle income countries. Regional differences are also large. Illiteracy rates are very high in Asia and Sub-Saharan Africa and relatively low in Latin America and the Caribbean.

Table 1. Illiteracy rates, by income level and regions, 1995 (percent)

Adult illiteracy rate	*Median*	*Unweighted average*	*Weighted average*
Income level			
Low income	48.0	46.9	34.4
Low income excluding China and India	48.3	47.5	45.6
Lower middle income	16.6	21.3	20.3
Upper middle income	16.6	15.8	14.5
Region			
South Asia	62.1	53.8	50.6
Sub-Saharan Africa	44.4	46.9	44.0
Middle East and North Africa	33.3	33.2	38.7
East Asia and the Pacific	16.5	17.5	16.9
Latin America and the Caribbean	10.2	15.7	13.4

Note: The unweighted average is the sum of illiteracy rates across countries in a given group divided by the number of countries in the group. The weighted average weights each country by the size of its population over the age of 15.
Source: EdStats, World Bank.

Regions with higher illiteracy rates tend to have lower GNP per capita and higher pupil-teacher ratios than regions with lower illiteracy rates (table 2). In Sub-Saharan Africa, for example, both GNP and literacy rates are much lower in Francophone countries than in Anglophone countries (table 3). Other indicators, such as enrollment rates and pupil-teacher ratios, are also weaker in Francophone countries.

Table 2. Selected social and macroeconomic indicators by region, 1995

Indicator	East Asia and Pacific	Eastern and Central Europe	Latin America and the Caribbean	Middle East and North Africa	South Asia	Sub-Saharan Africa
Adult illiteracy rate (percent)	16.9	–	13.4	38.7	50.6	44.0
GNP per capita (US$)	807.8	–	3,419.8	–	354.1	485.6
Gross enrollment ratio, primary (percent)	115.4	99.6	111.5	96.5	99.0	74.6
Pupil-teacher ratio, primary	24.2	20.0	24.5	27.8	62.7	40.6

Note: – Not available.
Source: EdStats, World Bank.

Table 3. Selected social and macroeconomic indicators for Sub-Saharan Africa

Indicator	*Anglophone countries*	*Francophone countries*
Adult illiteracy rate (percent)	36.1	51.9
GNP per capita (US$)	675.7	333.4
Gross enrollment ratio, primary (percent)	88.8	64.4
Pupils – teacher ratio, primary	37.8	47.3

Source: EdStats, World Bank.

Various measures of literacy improved over the sample period, 1975-85 to 1986-95 (table 4). The literacy rate rose significantly, increasing from 60.4 percent to 71.4 percent worldwide. Public expenditure on education as a fraction of total GDP increased, and the average number of years of schooling rose. Secondary school enrollment improved more than primary school enrollment, albeit starting at a much lower level. The pupil-teacher ratio, a proxy for quality of schooling, increased slightly. Health-related variables improved, the fertility rate fell, and life expectancy at birth increased.

Table 4. Average, lowest, and highest values for selected social indicators

Indicator	*Average*	*Lowest*	*Highest*	*Average*	*Lowest*	*Highest*
Literacy (percent)	60.4	13.3 Sierra Leone	99.5 United States	71.4	18.7 Burkina Faso	99.0 Georgia
Per capita GDP (US$)	3,204.4	98.8 Somalia	25,413.8 United Arab Emirates	4,318.5	107.0 Mozambique	33,088.0 Switzerland
Enrollment rate	91.0	16.0 Bhutan	144.0 Comoros	93.5	28.5 Ethiopia	142.0 Gabon
Education expenditure/ GDP (percent)	4.5	0.8 Somalia	9.8 Guyana	4.7	0.8 Nigeria	10.2 Tajikistan
Years of schooling	4.9	0.4 Guinea-Bissau	11.9 New Zealand	5.4	0.7 Guinea-Bissau	11.7 United States
Pupils per teacher	30.9	7.8 Norway	72.2 Mozambique	29.0	6.1 Norway	74.9 Cayman Islands
Fertility rate (births per woman)	4.5	1.4 Channel Islands	10.0 Oman	3.9	1.3 Hong Kong	7.5 Niger
Life expectancy (years)	61.2	35.3 Sierra Leone	76.0 Sweden	64.6	37.0 Sierra Leone	79.0 Japan
Urban/total population (percent)	12.5	Less than 1.0 Several	100.0 Singapore	14.3	0.0 Several	100.0 Singapore
Agriculture/ GDP (percent)	21.5	0.3 Kuwait	68.9 Guinea	20.8	0.3 Singapore	65.0 Somalia

Source: World Development Indicators, World Bank (1997).

4. Model and Methodology

4.1 The Model

We use regression analysis to analyze literacy, conditional on country specific characteristics. The model is based on the following equation:

$$y_i = \sum E_i \beta + \sum S_i \delta + \sum H_i \varphi + \sum Q_i \eta + \sum R_i \rho + \sum I_i \mu$$

The dependent variable, y_i, is the adult literacy rate in country *i*. The vector of explanatory variables consists of both demand and supply factors. We organize the explanatory variables into six groups: *E* represents economic development variables (per capita GDP, the share of agriculture value added in GDP, the share of the population living in urban areas). *S* represents the quality of education supplied (average public expenditure on education as a percentage of GNP, the average ratio of pupils to teachers in primary school). *H* represents health status (average fertility rate, average life expectancy at birth). *Q* represents the quality of education demanded (primary school enrollment rate, the average years of schooling of adults). *I* represents institutional variables—a combined measure consisting of riots (violent demonstrations or clashes of more than 100 people, wars on national territory, or civil wars) and a measure of democracy (Gastil's political rights)**.** *R* represents regional variables for Africa, Central and South America, Europe other than OECD countries, Oceania and Asia (other than OECD countries), and OECD countries. The vectors of parameters β, δ, φ, η, ρ, and μ reveal the marginal impact of each of the explanatory variables on literacy.

4.2 The Data

The data set was constructed from four different sources: World Development Indicators 1997, EdStats, the Penn-World Tables (Mark 5.6), and Easterly and Levine (1997) (see appendix A). The explanatory variables are averages for the 1975-85 period (the exceptions are GDP, for which data are from 1974, and the institutional variables, for which

data are from the decade of the 1970s). The dependent variable is the average literacy rate for 1986-95. The entire sample consists of 180 countries, but the actual number of countries used for a given model specification depends on the availability of explanatory variables: If one or more explanatory variables is missing for a given country, the country is dropped in the regression. Data constraints reduce the size of the sample to 57 countries in the initial specification and 68 countries in the final specification (the countries used in the initial specification are shown in appendix B). Natural logarithms are used for all data series except the urban population share.

4.3 Methodological Issues

This study (and all studies based on cross-country regressions) raises important methodological issues concerning aggregation; sampling; interpretation of coefficients; causality; measurement (including measurement error, index problems, and data availability); and the robustness of the results. (For an excellent discussion of these issues, see Levine and Renelt 1991a).

Questions about aggregation concern the appropriate unit of measurement. Are countries the appropriate unit to analyze, or should the analysis be conducted at a more disaggregated level? Variables measured at the aggregate level, such as income, health, and fertility, capture both components that are specific to a particular country and components that are part of worldwide trends. This is true for all macro analyses and hard to disentangle.

Sampling questions concern whether data are sampled from a single population—an implicit assumption underlying regression analysis. If all countries do not belong to the same underlying distribution of countries, some or all of the diagnostic tests would be violated and one or more countries would appear as outliers.

Questions about the interpretation of coefficients arise both because the regression coefficients represent intercountry averages and because the coefficients are not structural parameters but partial correlations between a given regressor and the regressand. As Levine and Renelt (1991a) note, "cross-country regressions may best be viewed as establishing patterns of correlation". We interpret the estimated parameters within a human capital framework and take note of Levine and Renelt's caveat.

We tackle the problem of causality by regressing average literacy rates on lagged average values of the regressors (that is, literacy in the second decade is determined by initial income and other regressors during the first decade). This suggests that the causality goes from the right-hand side to the left-hand side and not vice versa.

Measurement error may bias the parameter estimates. While it seems intuitive that measurement errors are related to factors such as country size, economic structure, and political instability, it is hard to do anything about it.

Levine and Renelt (1991b) warn that results should be interpreted with caution because they may not be very robust across different samples of countries and/or different time

periods. We try to deal with this problem by establishing that omitted observations and variable biases do not affect our results.

5. The Determinants of Literacy

5.1 Empirical Results

Multivariate regression techniques are used to identify the determinants of literacy rates. In addition to the variables already mentioned, a squared initial income variable is included in the regression to capture nonlinearities.

The first regression model reveals severe heteroscedasticity in the residuals. The problem is solved by accounting for regional differences in data (table 5). The test of homoscedastic residuals can not be rejected in the general regression model and the Ramsey reset test reveals no significant functional form misspecification of the model, suggesting that the model is well specified.

The model explains 84 percent of the variation in literacy across countries, with primary school enrollment, years of schooling, and life expectancy showing up as the most important explanatory variables. All three variables are statistically significantly different from zero and have the expected positive impact on literacy. Both an increase in health status (proxied by life expectancy at birth) and an increase in the number of years of schooling completed are unambiguously associated with increases in literacy.

TABLE 5. **Model Estimation Specification (Estimation of Literacy)**

Variable	*Coefficient*	*Standard error*	t-*statistic*	*P*>|*t*|
Enrollment	0.273	0.111	2.453	0.019
Years of school	0.367	0.075	4.880	0.000
Life expectancy	0.759	0.321	2.365	0.023
GDP	-0.767	0.582	-1.318	0.195
GDP squared	0.041	0.039	1.062	0.294
Agriculture/GDP	-0.088	0.062	-1.430	0.160
Fertility rate	0.059	0.122	0.482	0.632
Pupil/teacher	-0.096	0.102	-0.943	0.351
Educational expense/GDP	-0.019	0.082	-0.235	0.815
Urban/total population	-0.002	0.002	-0.964	0.341
Democracy	0.017	0.015	1.151	0.257
War and riots	0.025	0.056	0.445	0.658
Africa	-0.021	0.144	-0.142	0.887
Europe	0.100	0.156	0.642	0.525
Central/South America	0.033	0.119	0.278	0.783
Oceania and Asia	-0.107	0.133	-0.809	0.423
Constant	3.348	2.522	1.327	0.192

Note: $N = 57$. $F(16, 40) = 19.09$ Prob > $F = 0.000$. $R^2 = 0.88$, adjusted $R^2 = 0.84$. Cook-Weisberg test for heteroscedasticity: $\chi^2(1) = 5.10$ Prob > $\chi^2 = 0.024$. Ramsey RESET test for functional form: $F(3, 37) = 2.45$ Prob > $F = 0.079$. For definitions of variables, see appendix A.

The next step in the analysis is to reduce the general model to a more parsimonious specification. To do so, we perform sensitivity analysis by stepwise elimination of all combinations of one, two, three, four, and all five of the remaining explanatory variables in the regression (see appendix B). That analysis confirms that enrollment, years of schooling, and life expectancy are the key explanatory variables associated with literacy. Public expenditure on education proves important for preventing heteroscedasticity, but it is not statistically significant. We eliminate the share of urban population (urban/total population) and the war and riot variables, which are not statistically important in the regression analysis. The reduced model appears to be well specified, as neither the test of homoscedastic residuals nor the Ramsey reset test can be rejected (table 6).

Table 6. Reduced Model Specification (Estimation of Literacy)

Variable	*Coefficient*	*Standard error*	t-*statistic*	*P*>|*t*|
Enrollment	0.273	0.107	2.546	0.014
Years of school	0.3694	0.072	5.165	0.000
Life expectancy	0.763	0.301	2.539	0.014
GDP	-0.758	0.539	-1.405	0.166
GDP squared	0.042	0.036	1.155	0.254
Agriculture/GDP	-0.050	0.051	-0.987	0.328
Fertility rate	0.068	0.102	0.663	0.511
Pupil/teacher	-0.076	0.095	-0.793	0.431
Educational expenditure /GDP	-0.012	0.060	-0.192	0.848
Democracy	0.010	0.013	0.766	0.447
Africa	0.003	0.138	0.020	0.984
Europe	0.087	0.153	0.567	0.573
Central/South America	0.054	0.113	0.484	0.631
Oceania and Asia	-0.087	0.124	-0.703	0.485
Constant	3.029	2.342	1.294	0.202

Note: $N = 64$. $F(14, 49) = 23.16$ Prob > $F = 0.000$. $R^2 = 0.869$, adjusted $R^2 = 0.831$. Cook-Weisberg test for heteroscedasticity:$\chi^2(1) = 6.03$ Prob > $\chi^2 = 0.014$. Ramsey RESET test for functional form: $F(3, 46) = 2.46$ Prob > $F = 0.074$.

The final model (table 7), which excludes the statistically insignificant variables of agriculture's share of GDP, fertility, and pupil to teacher ratio, explains 85 percent of the variation of literacy rates around the world. It appears to be well specified, as neither the test of homoscedastic residuals nor the Ramsey reset test is rejected.

The model indicates that formal education has a significant impact on the literacy rate. A 10-percentage-point increase in the primary school enrollment rate is associated with a 2.6 percentage point increase in the literacy rate. Years of schooling appear to have an even greater impact, with a 10 percentage point increase in adults' years of schooling being associated with a 3.7 percentage point increase in the literacy rate.

Table 7.Final model specification (Estimation of Literacy)

Variable	*Coefficient*	*Standard error*	t-*statistic*	*P*>|*t*|
Enrollment	0.261	0.094	2.792	0.007
Years of school	0.368	0.066	5.590	0.000
Life expectancy	0.779	0.258	3.014	0.004
GDP	-0.775	0.483	-1.606	0.114
GDP squared	0.046	0.032	1.456	0.151
Educational expenditure/GDP	0.001	0.053	0.024	0.981
Democracy	0.008	0.011	0.752	0.455
Africa	0.063	0.113	0.552	0.583
Europe	0.062	0.136	0.449	0.655
Central/South America	0.060	0.090	0.664	0.510
Oceania and Asia	-0.039	0.100	-0.386	0.701
Constant	2.521	2.050	1.230	0.224

Note: $N = 68$. $F(11, 56) = 31.77$ Prob > $F = 0.000$. $R^2 = 0.862$, adjusted $R^2 = 0.835$. Cook-Weisberg test for heteroscedasticity: $\chi^2(1) = 4.84$ Prob > $\chi^2 = 0.028$. Ramsey RESET test for functional form: $F(3, 53) = 3.05$ Prob > $F = 0.036$.

The health status of the population (as measured by life expectancy) is positively correlated with literacy, with a one-year increase in life expectancy at birth associated with a 0.8 percentage point increase in the literacy rate. The quality of education, as measured by public expenditure on education, and the institutional variable are only weakly associated with literacy.

The relationship between income variables and literacy is nonlinear and only marginally statistically significant. (The variables are included to prevent heteroscedasticity and misspecification of the functional form.) At very low income levels, income and literacy are negatively associated. Once per capita income reaches a threshold of about $2,200 a year, the effect becomes positive.

Finally, African countries do not have a significantly higher literacy rate when controlling for other factors. Geographical location explains none of the variation in

worldwide literacy rates. (The regional dummy variables are included only to prevent heteroscedasticity.) The finding could mean that, controlling for the other variables included in the model, location has no effect on literacy. Alternatively, it could indicate that all of the region-specific variables that affect literacy are included in the model.

5.2 Omitted Observation and Variable Biases

The number of observations and variables in the regressions varies, because not all data series are available for all countries. To avoid omitted observation or variable biases, we specified two models based on the same sample of countries to observe whether the significance and/or magnitudes of estimated parameters were affected (tables 8 and 9).

The findings reveal that all results are robust to changes in the number of observations. No statistically significant variable becomes statistically insignificant with the change in the model. There are no sign changes, and the models remain well specified. These results suggest that our model is robust and that no omitted variable or observation biases exist.

Table 8. Model 2

Variable	*Coefficient*	*Standard error*	t-*statistic*	*P*>|*t*|
Enrollment	0.288	0.109	2.638	0.012
Years of school	0.364	0.073	4.979	0.000
Life expectancy	0.800	0.311	2.572	0.014
GDP	-0.866	0.543	-1.596	0.118
GDP squared	0.048	0.036	1.307	0.198
Agriculture/GDP	-0.064	0.056	-1.149	0.257
Fertility rate	0.103	0.113	0.914	0.366
Pupil/teacher	-0.106	0.098	-1.077	0.287
Education expenditure./GDP	-0.021	0.077	-0.276	0.784
Democracy	0.010	0.013	0.764	0.449
Africa	-0.015	0.142	-0.102	0.919
Europe	0.088	0.153	0.573	0.570
Central/South America	0.028	0.118	0.238	0.813
Oceania and Asia	-0.117	0.130	-0.897	0.375
Constant	3.426	2.366	1.448	0.155

Note: $N = 57$. $F(14, 42) = 22.24$ Prob > $F = 0.000$. $R^2 = 0.881$, adjusted $R^2 = 0.842$. Cook-Weisberg test for heteroscedasticity:χ^2 (1) = 5.63 Prob > $\chi^2 = 0.018$. Ramsey RESET test for functional form: $F(3, 46) = 2.55$ Prob > $F = 0.067$.

Table 9. Model 3

Variable	*Coefficient*	*Standard error*	t-*statistic*	*P*>|*t*|
Enrollment	0.250	0.101	2.478	0.017
Years of school	0.385	0.071	5.455	0.000
Life expectancy	0.788	0.276	2.861	0.006
GDP	-0.818	0.514	-1.591	0.119
GDP squared	0.048	0.034	1.421	0.162
Education expenditure/GDP	0.008	0.066	0.128	0.899
Democracy	0.012	0.012	0.939	0.353
Africa	0.071	0.130	0.545	0.589
Europe	0.080	0.148	0.544	0.589
Central/South America	0.091	0.104	0.873	0.387
Oceania and Asia	-0.032	0.116	-0.279	0.782
Constant	2.702	2.197	1.230	0.225

Note: $N = 57$. $F(1, 45) = 28.52$ Prob > $F = 0.000$. $R^2 = 0.875$, adjusted $R^2 = 0.844$. Cook-Weisberg test for heteroscedasticity: $\chi^2(1) = 4.14$ Prob > $\chi^2 = 0.042$. Ramsey RESET test for functional form: $F(3, 53) = 2.91$ Prob > $F = 0.043$

6. Summary

Formal education has an important impact on literacy rates, with both school enrollment rates and average years of schooling of adults being positively associated with increased literacy. A 10-percentage-point increase in the primary school enrollment rate is associated with a 2.6 percentage point increase in the literacy rate. A 10-percentage-point increase in adult's years of schooling is associated with a 3.7-percentage-point increase in the literacy rate.

Health is also an important factor affecting literacy. When life expectancy at birth increases one year, the literacy rate increases by 0.8 percentage points.

The effect of income on literacy is nonlinear. Income is negatively associated with literacy until a threshold of about $2200 a year, after which the effect becomes positive. These findings suggest that countries must reach a certain level of economic development before education pays off. Holding all other factors constant, location explains none of the variation in worldwide literacy. Finally, the quality of education and institutions are not statistically significant determinants of literacy. Finally, African countries do not have a significantly higher illiteracy rate when controlling for other factors.

References

Blunch, N.H. and D. Verner. 1999. "Determinants of Income and Literacy in Ghana", Mimeo, World Bank.

Cochrane, S. H. 1979. *Fertility and Education: What Do We Really Know?* Baltimore: Johns Hopkins University Press.

Coy, J. J. 1988. "Adult Literacy from a Sociocultural Perspective." Paper presented at the 26th annual meeting of the Florida Reading Association.

Easterly, W., and R. Leaven. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112 (4): 1203-50.

Hass, R. 1996. "Passing on the Gift: The Importance of Literacy in America." *Teachers & Writers* 27 (4).

Levine, R. and D. Renelt. 1991a. "Cross-Country Studies of Growth and Policy. Methodological, Conceptual, and Statistical Problems." Working Paper 608. Macroeconomic Adjustment and Growth Division, Country Economics Department, World Bank, Washington, D.C.

———. 1991b. "A Sensitivity Analysis of Cross-Country Growth Regressions." Working Paper Series 609. Macroeconomic Adjustment and Growth Division, Country Economics Department, World Bank, Washington, D.C.

Neuman, Shoshana and Avi Weiss. 1995. "On the Effects of Schooling Vintage on Experience- Earnings Profiles: Theory and Evidence". European Economic Review 39 (5).

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 1993. *World Education Report*. Paris.

Verner, Dorte. 1995. "Can the Variables in an Extended Solow Model be Treated as Exogenous?" ECO No.95/20, EUI, Italy.

World Bank. 1991. *World Development Report.* New York: Oxford University Press.

———. 1997. *World Development Indicators.* Washington, D.C.

———. *EdStats.* (Various years). Washington, D.C.

Appendix A

Variable Definitions and Description of Sample

Data from EDStats, World Bank

Enrollment: Average gross enrollment in primary school (log), where gross enrollment in primary school is the percentage of school-age children attending school. Many but not all countries consider primary school age to be 6-11 years. For some countries with universal primary education, gross enrollment ratios may exceed 100 percent because some pupils are younger or older than the countries' standard primary school age.

Literacy: Average adult literacy rate (log)

Pupil/teacher: Average pupil-teacher ratio in primary school (log)

Education expenditure/GDP: Average public expenditure on education as percentage of GNP (log)

Years of school: Average years of schooling of adults (log)

Data from *World Development Indicators 1997*, World Bank

Agriculture/GDP: Average value added in agriculture as a percentage of GDP (log)

Fertility rate: Average total fertility rate, where total fertility rate is the number of births per woman of childbearing age (15-49) (log). The total fertility rate represents the number of children that would be born by a woman if she were to live to the end of her child bearing age and bear children at each age in accordance with prevailing age specific fertility rates.

Life expectancy: Average life expectancy at birth, total number of years (log)

Urban/total population: Average urban population, where urban population is calculated as a share of the total population. Estimates are based on different national definitions of what is urban; cross-country comparison should thus be made with caution.

Data from Penn-World Tables (Mark 5.6)

GDP: Real per capita gross domestic product in 1974 at 1985 international prices, calculated using a chain index (log)

GDP squared: GDP squared

Data from Easterly and Levine (1997)

Riots: 1 if violent demonstration or clash involving more than 100 people occurred during the decade, 0 otherwise

War: 1 if war took place on national territory during the decade, 0 otherwise

Warc: 1 if there was a civil war during the decade, 0 otherwise.

War and riots: 1 if either *riots*, *war,* or *warciv* was 1 during the decade, 0 otherwise

Democracy: Measure of democracy (Gastil's political rights) for the 1970s

Regional Dummies

Africa: 1 for Africa, 0 otherwise

Europe: 1 for Europe (other than OECD countries), 0 otherwise

Central/South America: 1 for Central and South America, 0 otherwise

Oceania and Asia: 1 for Oceania and Asia (other than OECD countries), 0 otherwise

Description of Sample

Statistics on the following countries or territories were not available in EdStats and were thus omitted from the data set: American Samoa, Andorra, Aruba, Bermuda, Bosnia and Herzegovina, Brunei, Cayman Islands, Chad, Channel Islands, Faeroe Islands, French Guyana, French Polynesia, Greenland, Guadeloupe, Guam, Iceland, Isle of Man, Democratic Republic of Korea, Liechtenstein, Luxembourg, Macao, Martinique, Mayette, Micronesia, Monaco, Netherlands Antilles, New Caledonia, Northern Mariana Islands, Reunion, Virgin Islands**,** and Federal Republic of Yugoslavia. GDP data on Germany from the Penn-World Tables refer to West-Germany.

Appendix B

Sensitivity Analysis

Table A1. Countries used in the sensitivity analysis

1	Togo	20	Paraguay	39	Italy
2	Nepal	21	Cameroon	40	Gambia
3	Egypt	22	Peru	41	Philippines
4	Pakistan	23	Dominican Republic	42	Jamaica
5	India	24	Jordan	43	Bolivia
6	Sierra Leone	25	Mauritius	44	Singapore
7	Chile	26	Sri Lanka	45	Turkey
8	Argentina	27	Mexico	46	Mali
9	Honduras	28	Ghana	47	Malaysia
10	Algeria	29	Kenya	48	Indonesia
11	Greece	30	Syria	49	Congo
12	Sudan	31	Nicaragua	50	Brazil
13	Spain	32	Korea	51	Mozambique
14	Bangladesh	33	Venezuela	52	Thailand
15	Trinidad and Tobago	34	Botswana	53	Uganda
16	China	35	Malawi	54	El Salvador
17	Lesotho	36	Uruguay	55	Rwanda
18	Niger	37	Zambia	56	Zimbabwe
19	Tunisia	38	Tanzania	57	Myanmar

Table A2. Results of sensitivity analysis

Regression	*Enrollment, primary school*	*Years of schooling*	*Life expectancy*	*Log GDP 1974*	*Log GDP 1974 squared*	*Agriculture/GDP*	*Fertility rate*	*Pupil/ teacher ratio*	*Democracy 1970*	*Public education expenditure/GDP*	*Test for heteroscedasticity*
0	1+	1+	1+	3-	4+			IS	IS	IS	0.014
1	1+	1+	1+	2-	3+		--IS	IS	IS	IS	0.0186
2	1+	1+	1+	4-	IS	IS	--	IS	IS	IS	0.0177
3	1+	1+	1+	4-	4+	IS	IS	--	IS	IS	0.0245
4	1+	1+	1+	2-	3+	IS	4+	IS	--	IS	0.0095
5	1+	1+	2+	3-	4+	IS	IS	IS	IS	--	0.0001
6	1+	1+	1+	2-	3+	--	--	IS	IS	IS	0.021
7	1+	1+	1+	4	IS	IS	--	--	IS	IS	0.027
8	1+	1+	1+	3-	3+	IS	IS	--	--	IS	0.018
9	1+	1+	2+	3-	4+	IS	IS	1-	--	--	0.0001
10	1+	1+	1+	3-	4+	--	IS	--	IS	IS	0.026
11	1+	1+	1+	2-	2+	--	IS	IS	--	IS	0.016
12	1+	1+	2+	3-	3+	--	IS	2-	IS	--	0.0001
13	1+	1+	1+	3-	3	IS	--	IS	--	IS	0.012
14	1+	1+	3+	4-	IS	IS	--	2-	IS	--	0.0001
15	1+	1+	2+	IS	IS	IS	IS	--	IS	--	0.0000
16	1+	1+	1+	3-	3+	--	--	--	IS	IS	0.028
17	1+	1+	1+	3-	4+	IS	--	--	--	IS	0.019
18	1+	1+	2+	3-	4+	IS	IS	--	--	--	0.0000
19	1+	1+	1+	2-	2+	--	IS	--	--	IS	0.023
20	1+	1+	2+	3-	3+	--	IS	2-	--	--	0.0000
21	1+	1+	1+	2-	2+	--	--	IS	--	IS	0.017
22	1+	1+	2+	3-	3+	--	--	2-	IS	--	0.0001
23	1+	1+	3+	4-	4+	IS	--	2-	--	--	0.0001
24	1+	1+	2+	4-	IS	IS	--	--	IS	--	0.0000
25	1+	1+	2+	4-	4+	--	IS	--	IS	--	0.0000
26	1+	1+	1+	1-	3+	--	--	--	--	IS	0.0225
27	1+	1+	1+	4-	IS	IS	--	--	--	--	0.0000
28	1+	1+	1+	4-	4+	--	IS	--	--	--	0.0000
29	1+	1+	1+	3-	3+	--	--	--	IS	--	0.0000
30	1+	1+	2+	3-	3+	--	--	2-	--	--	0.0000
31	1+	1+	1+	3-	3+	--	--	--	--	-- IS	0.0000

Note: 1 = *p*-value of 0-5 percent; 2 = *p*-value >5 percent to 10 percent; 3 = *p*-value >10 percent to 20 percent; 4 = *p*-value >20 percent to 30 percent. IS = insignificant (*p*-value > 30 percent); -- =omitted in the regression; + =positive coefficient; - = negative coefficient. Last column is *p*-value from test for heteroscedasticity of the residuals.